# Aspects probabilistes de l'informatique

Anne Bouillard

Notes de cours pour l'année 2017-2018

## Table des matières

## Bibliographie

Les sources principales sur lequel se basens ces notes de cours sont les suivantes.

[1] Michael Mitzenmacher et Eli Upfal. *Probability and Computing : Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.

[2] Noda Alon et Joel H. Spencer, *The probabilistic method*, John Wiley & Sons, 2000.

[3] Moez Draief et Laurent Massoulié, *Epidemics and Rumours in Complex Networks*, London Mathematical Society Lecture Note Series, 2010.

[4] Pierre Brémaud. *Initiation aux probabilités et aux chaînes de Markov*, Springer, 2009.

[5] Pierre Brémaud. *Markov Chains : Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer, 1999.

[6] Olle Häggström. *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press. 2002.

# 1 Rappels de probabilités

## 1.1 Événements et probabilités

**Exemple (lancer de dés).** On dit « une chance sur 6 d'obtenir un 3 » ou « une chance sur deux d'obtenir un chiffre pair ». Informellement, « obtenir une 3 » et « obtenir un chiffre pair » sont des *événements* et leurs probabilités respectives sont 1/6 et 1/2.

L'exercice suivant montre qu'une définition informelle peut donner lieu à des ambiguïtés.

**Exercice 1** *Tirer une corde au hasard*
Quelle est la probabilité qu'une corde d'un cercle *choisie au hasard* soit plus grande que le côté du triangle équilatéral inscrit dans le même cercle ?

Trouver plusieurs manières de « choisir au hasard » qui donnent des probabilités différentes.

Il est donc nécessaire de définir rigoureusement la notion de « choisie au hasard ». Pour ce faire, on introduit la notion d'espace probabiliste.

### 1.1.1 Tribus et événements

On se donne
- $\Omega$ un ensemble qui décrit toutes les possibilités d'une expérience. Par exemple, pour un dé, on a $\Omega = \{1, 2, 3, 4, 5, 6\}$. Les éléments de $\Omega$ sont appelés les *épreuves* ou les *réalisations* ;
- intuitivement, un *événement* est un sous-ensemble de $\Omega$. Pour un dé par exemple, « obtenir un 3 » correspond à l'événement $\{3\}$ et « obtenir un chiffre pair » correspond à l'événement $\{2, 4, 6\}$.

Définissons maintenant ces concepts de manière formelle.

**Définition 1.** *Une* tribu *sur $\Omega$ est une famille $\mathcal{F}$ de sous-ensembles de $\Omega$ telle que*

($\alpha$) *$\Omega$ et $\emptyset$ sont dans $\mathcal{F}$ ($\Omega, \emptyset \in \mathcal{F}$).*

($\beta$) *si $A \in \mathcal{F}$, alors $A^c \triangleq \bar{A} \triangleq \Omega \setminus A \in \mathcal{F}$ (stabilité par complémentaire).*

($\gamma$) *si $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ alors $\bigcup_{n=0}^{\infty} A_n \in \mathcal{F}$ (stabilité par union dénombrable).*

La tribu grossière est la plus petite tribu sur $\Omega$ et c'est $\{\Omega, \emptyset\}$. La tribu fine est la plus grosse tribu sur $\Omega$ et c'est $\{A \mid A \subseteq \Omega\}$.

Les événements doivent former une tribu. En général, dans le cadre de ce cours, on considérera la tribu fine, car l'on se place dans le cas où $\Omega$ est au plus dénombrable, et dans ce cas, $\mathcal{P}(\Omega)$ est engendré par les singletons.

Notons que ($\beta$)+($\gamma$) entraîne la stabilité par intersection dénombrable : $\bigcap_{n\in\mathbb{N}} A_n = (\bigcup_{n\in\mathbb{N}} A_n^c)^c$.

**Exemple (événements plus complexes : petite incursion dans les espaces non dénombrables).** Soit $\Omega = \{0, 1\}^{\mathbb{N}}$. Les épreuves sont les éléments $\omega = (\omega_0, \omega_1, \omega_2, \ldots)$, $\omega_i \in \{0, 1\}$. Soit $A = \{\omega \mid w_i \in A_i, i \leq k\}$ avec $A_i \subseteq \{0, 1\}$. En d'autres termes, l'espace des épreuves peut se voir comme une suite infinie de lancers de pièce et l'événement $A$ concerne uniquement les $k$ premiers lancers. Les événements de type $A$ ne sont pas stables par union dénombrable. Par exemple, l'événement $\{\omega \mid \lim_{n\to\infty} \omega_n = 0\}$ représente les épreuves ayant un nombre fini de 1 mais appartient à la tribu engendrée par ces événements :

$$A = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \{\omega \mid \omega_m = 0\}.$$

### 1.1.2 Espace de probabilités, axiomes des probabilités

**Définition 2.** *Soit $\Omega$ un espace d'épreuves et $\mathcal{F}$ une tribu sur $\Omega$. Une* probabilité *sur $(\Omega, \mathcal{F})$ est une application $\mathbf{P} : \mathcal{F} \to [0,1]$ telle que*

($\alpha$) $\mathbf{P}(\Omega) = 1$ ;

($\beta$) ($\sigma$-additivité) *si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements* deux à deux disjoints *alors*

$$\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbf{P}(A_n).$$

On appelle $(\Omega, \mathcal{F}, \mathbf{P})$ un *espace de probabilités.*

Soit $A$ un événement. Si $\mathbf{P}(A) = 1$, alors on dit que $A$ est presque sûr. Si $\mathbf{P}(A) = 0$, alors on dit que $A$ est presque impossible.

**Exemple (construction de probabilité).** On interprète 0 comme pile et 1 comme face. On reprend l'exemple $\Omega = \{0,1\}^{\mathbb{N}}$ et on note $E_n = \{w \mid w_n = 1\}$ l'événement « le $i$-ème lancer est face ».

On veut trouver une probabilité telle que

— $\mathbf{P}(E_n) = p$, et donc $\mathbf{P}(E_n^c) = 1 - p$ et

— on a « indépendance des lancers » : $\mathbf{P}(\cap_{i\leq n} E_i^{(c)}) = \prod_{i\leq n} \mathbf{P}(E_i^{(c)})$, où $E_i^{(c)}$ représente soit $E_i$, soit $E_i^c$.

Alors, on doit nécessairement choisir

$$\mathbf{P}(\cap_{i\leq n} E_i^{(c)}) = p^{\sum_{i=0}^n a_i}(1-p)^{n-\sum_{i=0}^n a_i}$$

où $a_i = 1$ si $E_i$ apparaît ou $= 0$ si c'est $E_i^c$ qui est dans la formule.

On peut montrer (hors programme) que ceci suffit à bien définir une probabilité.

**Proposition 1.** *Soient $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilités et $A$, $B$ et $A_n$, $n \in \mathbb{N}$ des événements (de $\mathcal{F}$).*

1. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.
2. $\mathbf{P}(\emptyset) = 0$.
3. $A \subseteq B \Rightarrow \mathbf{P}(A) \leq \mathbf{P}(B)$ (monotonie).
4. $\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} \mathbf{P}(A_n)$ (union-bound).
5. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

*Démonstration.*

1. $\mathbf{P}(A) + \mathbf{P}(A^c) = \mathbf{P}(\Omega) = 1$.
2. $\mathbf{P}(\emptyset) = \mathbf{P}(\Omega^c) = 1 - \mathbf{P}(\Omega) = 1 - 1 = 0$.
3. $\mathbf{P}(B) = \mathbf{P}(B \setminus A) + \mathbf{P}(A)$. Mais $\mathbf{P}(B \setminus A) \geq 0$ donc $\mathbf{P}(B) \geq \mathbf{P}(A)$.
4. Posons $B_n = A_n - \cup_{k<n} A_k \subseteq A_n$. Les $B_n$ sont deux à deux disjoints donc $\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \mathbf{P}(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbf{P}(B_n) \leq \sum_{n\in\mathbb{N}} \mathbf{P}(A_n)$.
5. $\mathbf{P}(A) = \mathbf{P}(A \setminus (A \cap B)) + \mathbf{P}(A \cap B)$. De même, $\mathbf{P}(B) = \mathbf{P}(B \setminus (A \cap B)) + \mathbf{P}(A \cap B)$ et $\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus (A \cap B)) + \mathbf{P}(B \setminus (A \cap B)) + \mathbf{P}(A \cap B)$. Au final, on a donc bien $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

□

**Théorème 1** (Continuité séquentielle)**.** *Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'événements ($\forall n \in \mathbb{N}$, $A_n \subseteq A_{n+1}$). Alors*

$$\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

**Continuité séquentielle**

*Démonstration.* Posons $B_n = A_n \setminus A_{n-1}$ (avec la convention $A_{-1} = \emptyset$). Alors $\cup_{n\in\mathbb{N}} A_n = \cup_{n\in\mathbb{N}} B_n$ et les $B_n$ sont deux à deux disjoints, et $A_n = \cup_{k\leq n} B_k$. Donc,

$$\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \mathbf{P}(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbf{P}(B_n) = \lim_{n\to\infty} \sum_{k\leq n} \mathbf{P}(B_k) = \lim_{n\to\infty} \mathbf{P}(\cup_{k\leq n} B_k) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

□

**Corollaire 1.** *Soit $(B_n)_{n\in\mathbb{N}}$ une suite décroissante d'événements ($\forall n \in \mathbb{N}$, $B_{n+1} \subseteq B_n$). Alors*

$$\mathbf{P}(\cap_{n\in\mathbb{N}} B_n) = \lim_{n\to\infty} \mathbf{P}(B_n).$$

*Démonstration.*

$$\mathbf{P}(\cap_{n\in\mathbb{N}} B_n) = 1 - \mathbf{P}(\overline{\cap_{n\in\mathbb{N}} B_n}) = 1 - \mathbf{P}(\cup_{n\in\mathbb{N}} B_n^c) = 1 - \lim_{n\to\infty} \mathbf{P}(B_n^c) = \lim_{n\to\infty} (1 - \mathbf{P}(B_n^c)) = \lim_{n\to\infty} \mathbf{P}(B_n).$$

□

### 1.1.3 Indépendance, probabilité conditionnelle

**Exemple (égalité de deux polynômes.).** Reprenons l'exemple de la vérification de l'égalité de deux polynômes présenté en introduction. Soient $F$ et $G$ deux polynôme degré $d$. La probabilité que l'algorithme échoue est $\mathbf{P}(\text{algorithme échoue}) \leq 1/100$.

Que faire si l'on veut une plus grande précision ?

1. Augmenter l'espace des entiers dans lequel on choisit la racine potentielle. Ce n'est pas satisfaisant car pour de grands entiers, cela mènerait à des problèmes de précision.
2. Répéter l'algorithme plusieurs fois. Si l'on a une sortie $r$ telle que $F(r) \neq G(r)$, alors on sait que $F \neq G$. On peut choisir avec ou sans remise des valeurs déjà tirées. La première solution ne dépend alors pas des résultats déjà obtenus alors que la seconde en dépend.

Avec remise, intuitivement, faire $k$ essais entraîne une probabilité de se tromper d'au plus $(1/100)^k$, sans remise, cette probabilité sera plus petite.

**Définition 3.** *Deux événements $A$ et $B$ sont indépendants si $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$. $(A_n)_{n\in\mathbb{N}}$ est une famille d'événements mutuellement indépendants si pour toute famille finie $(A_{i_0}, \ldots, A_{i_n})$, $\mathbf{P}(A_{i_0} \cap \cdots \cap A_{i_n}) = \mathbf{P}(A_{i_0}) \cdots \mathbf{P}(A_{i_n})$.*

**Indépendance d'évenements**

**Exercice 2**

$(A_n)_{n\in\mathbb{N}}$ est une famille d'événements mutuellement indépendants si et seulement si la famille $(B_n)_{n\in\mathbb{N}}$ telle que $\forall n \in \mathbb{N}$, $B_n = A_n$ ou $A_n^c$ l'est.

**Exercice 3**

L'indépendance mutuelle n'est pas l'indépendance deux à deux.

**Théorème 2.** *Soit $(C_n)_{n\in\mathbb{N}}$ une suite d'événements mutuellement indépendants. Alors*

$$\mathbf{P}(\cap_{n\in\mathbb{N}} C_n) = \prod_{n\in\mathbb{N}} \mathbf{P}(C_n).$$

*Démonstration.* On utilise la continuité séquentielle. Soit $B_n = \cap_{k\leq n} C_k$. Alors $(B_n)_{n\in\mathbb{N}}$ est une suite décroissante d'événements et d'une part,

$$\mathbf{P}(B_n) = \mathbf{P}(\cap_{k=0}^{n} C_k) = \prod_{k=0}^{n} \mathbf{P}(C_k),$$

et d'autre part,

$$\lim_{n\to\infty} \mathbf{P}(B_n) = \mathbf{P}(\cap_{n\in\mathbb{N}} B_n) = \mathbf{P}(\cap_{n\in\mathbb{N}} C_n) \text{ et } \lim_{n\to\infty} \prod_{k=0}^{n} \mathbf{P}(C_k) = \prod_{k=0}^{\infty} \mathbf{P}(C_k).$$

Le résultat s'en déduit immédiatement. □

Dans l'algorithme précédemment décrit, on tire à chaque fois une valeur au hasard entre 1 et 100d. Ces tirages sont effectués indépendamment. Soit $E_i$ l'événement « je tire une racine pour le $i$-ème tirage ». La probabilité que l'algorithme se trompe est alors $\mathbf{P}(\cap_{i=1}^{k} E_i) = \prod_{i=1}^{k} \mathbf{P}(E_i) \leq (1/100)^k$. L'erreur décroît donc de manière exponentielle.

**Probabilité conditionnelle** Supposons maintenant que l'on ne tire pas une valeur déjà tirée. Les événements $E_i$ (la $i$-ème valeur tirée est une racine) ne sont donc plus indépendants. On introduit la notion de *probabilité conditionnelle*.

**Définition 4.** *La probabilité d'un événement $A$ étant donné (ou conditionné à) un événement $B$ est définie par*

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

*Cette probabilité est bien définie uniquement si $\mathbf{P}(B) \neq 0$. Dans le cas contraire, on peut définir $\mathbf{P}(A \mid B)$ de manière arbitraire.*

**Remarque 1.** *Si deux événements sont indépendants, alors $\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A)\mathbf{P}(B)}{\mathbf{P}(B)} = \mathbf{P}(A)$.*

Les probabilités sans remise deviennent (notons que l'on omet le signe $\cap$ dans les formules, cette omission deviendra quasiment automatique dans la suite) :

$$\begin{aligned}
\mathbf{P}(E_1 \cap \cdots \cap E_k) &= \mathbf{P}(E_k \mid E_1 \cap \cdots \cap E_{k-1})\mathbf{P}(E_1 \cap \cdots \cap E_{k-1}) \\
&= \mathbf{P}(E_k \mid E_1 \cap \cdots \cap E_{k-1})\mathbf{P}(E_{k-1} \mid E_1 \cap \cdots \cap E_{k-2})\mathbf{P}(E_1 \cap \cdots \cap E_{k-2}) \\
&= \mathbf{P}(E_1)\mathbf{P}(E_2 \mid E_1)\mathbf{P}(E_3 \mid E_1, E_2) \cdots \mathbf{P}(E_k \mid E_1, \ldots, E_{k-1})
\end{aligned}$$

Or $\mathbf{P}(E_i \mid E_1 \cdots E_{i-1}) \leq \frac{d-(j-1)}{100d-(j-1)}$ (avec égalité si les $d$ racines sont distinctes), donc pour $k \leq d$,

$$\mathbf{P}(E_1 \cap \cdots \cap E_k) \leq \prod_{j=1}^{k} \frac{d-(j-1)}{100d-(j-1)} \leq \left(\frac{1}{100}\right)^k.$$

On obtient une erreur plus petite qu'avec remise, mais en pratique il est parfois plus judicieux algorithmiquement d'utiliser la méthode avec remise.

La formule utilisée est le théorème de Bayes :

**Théorème 3** (Loi de Bayes séquentielle)**.** *Soient $E_1, \ldots, E_n$ des événements. La probabilité de l'intersection de ces événements peut se calculer comme*

$$\mathbf{P}(E_1 \cap \cdots \cap E_k) = \mathbf{P}(E_1)\mathbf{P}(E_2 \mid E_1)\mathbf{P}(E_3 \mid E_1 \cap E_2) \cdots \mathbf{P}(E_k \mid E_1 \cap \cdots \cap E_{k-1}).$$

La preuve se fait par induction en utilisant le principe évoqué plus haut.

On peut aussi déduire de la définition des probabilités conditionnelles les lois (de Bayes) suivantes.

**Théorème 4** (Loi des probabilités totales)**.** *Soit $(E_1, \ldots, E_n)$ une partition de $\Omega$ (les événements $E_1, \ldots, E_n$ sont* 2 à 2 disjoints *tels que $\cup_{i=1}^n E_i = \Omega$). Alors pour tout événement $A$,*

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A \mid E_i)\mathbf{P}(E_i).$$

*Démonstration.*

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{P}(\cup_{i=1}^n (A \cap E_i)) \\ &= \sum_{i=1}^{n} \mathbf{P}(A \cap E_i) = \sum_{i=1}^{n} \mathbf{P}(A \mid E_i)\mathbf{P}(E_i). \end{aligned}$$

□

**Proposition 2** (Loi de rétrodiction)**.** *Pour tous événements $A$ et $B$ avec $\mathbf{P}(B) > 0$,*

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\mathbf{P}(B \mid A).$$

*Démonstration.*

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\mathbf{P}(B \mid A).$$

□

**Exercice 4**
On se donne trois pièces dont une et une seule est biaisée de telle sorte que $\mathbf{P}(\texttt{Face}) = 2/3$. On lance les pièces et on obtient `Face Face Pile`.

Quelle est la probabilité que la première pièce soit la pièce biaisée ?

## 1.2 Variables aléatoires

### 1.2.1 Variables aléatoires et distribution

**Définition 5.** *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soit $E$ un ensemble au plus dénombrable. Une fonction $X : \Omega \to E$ telle que pour tout $x \in E$, $\{\omega \mid X(\omega) = x\} \in \mathcal{F}$ est une* variable aléatoire (v.a.) discrète sur $E$.

On note dans la suite $\{X = x\}$ ou plus simplement $X = x$ l'événement $\{\omega \mid X(\omega) = x\}$ et $\{X \in A\}$ ou plus simplement $X \in A$ l'événement $\{\omega \mid X(\omega) \in A\}$.

Si $E \subseteq \mathbb{R}$, on parle de variable aléatoire réelle (v.a.r.).

**Définition 6.** *Soit $E$ un ensemble au plus dénombrable. Soit $\{p(x),\ x \in E\}$ une suite de réels tels que $\forall x \in E$, $0 \leq p(x) \leq 1$ et $\sum_{x\in E} p(x) = 1$. Cette suite est appelée* distribution de probabilité *sur $E$. Si une variable aléatoire (v.a.) $X$ vérifie $\mathbf{P}(X \in A) = \sum_{x\in A} p(x)$ pour tout $A \subseteq E$, on dit que $(p(x),\ x \in E)$ est la distribution (ou loi) de probabilité de $X$.*

On peut adapter les définitions vues pour les événements aux variables aléatoires :

**Définition 7.**
- *Deux variables aléatoires $X$ et $Y$ sont* indépendantes *si et seulement si pour tous $x$ et $y$, $\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y)$ ;*
- *Les variables aléatoires $X_1, \ldots, X_k$ sont* mutuellement indépendantes *si et seulement si $\forall I \subseteq \{1, \ldots, k\}$, $\forall (x_i)_{i\in I}$,*

$$\mathbf{P}(\bigcap_{i\in I}\{X_i = x_i\}) = \prod_{i\in I} \mathbf{P}(\{X_i = x_i\}) \ ;$$

- *les variables aléatoires $X_1, \ldots, X_k$ sont* indépendantes et identiquement distribuées (i.i.d.) *si et seulement si elles sont mutuellement indépendantes et de même distribution.*

**Proposition 3.** *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soient $X_1, \ldots, X_n$ des variables aléatoires à valeurs dans respectivement $E_1, \ldots, E_n$ des ensembles au plus dénombrables et $f : E_1 \times \cdots \times E_n \to F$ une fonction. Alors $f(X_1, \ldots, X_n)$ est une variable aléatoire.*

*Démonstration.* $X_i$ est une variable aléatoire, donc $\{\omega \mid X_i(\omega) = y_i\} \in \mathcal{F}$ pour tout $y_i \in E_i$. Or, pour tout $x \in F$,

$$\{\omega \mid f(X_1(\omega), \ldots, X_n(\omega)) = x\} = \bigcup_{\{y_i,\ldots,y_n \mid f(y_1,\ldots,y_n)=x\}} \bigcap_{i=1}^{n} \{\omega \mid X_i(\omega) = y_i\}$$

Donc $\{f(X_1, \ldots, X_n) = x\} \in \mathcal{F}$ par stabilité par intersection et union dénombrable. □

En particulier, si $X$ et $Y$ sont des variables aléatoires et $f$ une fonction, $X + Y$, $XY$, $f(X)$ en sont aussi.

**Proposition 4.** *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soient $X_1, \ldots, X_n$ des variables aléatoires mutuellement indépendantes à valeurs dans respectivement $E_1, \ldots, E_n$ des ensembles au plus dénombrables et $f_i : E_i \to F_i$ une fonction. Alors $f_1(X_1), \ldots, f_n(X_n)$ sont des variables aléatoires mutuellement indépendantes.*

*Démonstration.* Nous montrons le résultat pour $n = 2$ seulement, le cas général n'est qu'une généralisation d'écriture.

Pour tout $(j_1, j_2) \in F_1 \times F_2$,

$$\begin{aligned} p(f_1(X_1) = j_1, f_2(X_2) = j_2) &= \sum_{i_1\in E_1, f_1(i_1)=j_1} \sum_{i_2\in E_2, f_2(i_2)=j_2} \mathbf{P}(X_1 = i_1, X_2 = i_2) \\ &= \sum_{i_1\in E_1, f_1(i_1)=j_1} \mathbf{P}(X_1 = i_1) \sum_{i_2\in E_2, f_2(i_2)=j_2} \mathbf{P}(X_2 = i_2) \\ &= \mathbf{P}(f_1(X_1) = j_1)\mathbf{P}(f_2(X_2) = j_2). \end{aligned}$$

□

**Exemples de distributions**

1. **Loi constante égale à $a$ : $\mathbf{P}(X = a) = 1$.**
2. **Loi de Bernoulli :**

$$X \sim \mathcal{B}\text{er}(p) \text{ si } \mathbf{P}(X = 1) = p \text{ et } \mathbf{P}(X = 0) = 1 - p.$$

*Interprétation :* pile ou face biaisé. Un exemple fondamental de v.a. de ce type de distribution est la fonction caractéristique d'un événement : pour $A \in \mathcal{F}$, $\mathbf{1}_A$ définie telle que $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon est une v.a. de loi Bernoulli de paramètre $\mathbf{P}(A)$.

3. **Loi binomiale de paramètres $n$ et $p$ :**

$$X \sim \mathcal{B}\text{in}(n, p) \text{ si } \mathbf{P}(X = j) = \binom{n}{j} p^j (1-p)^{n-j}.$$

*Interprétation :* $n$ lancers d'une pièce qui donne face avec probabilité $p$. $p(j)$ est la probabilité d'obtenir $j$ fois face exactement parmi $n$ lancers indépendants. $X = \sum_{i=1}^{n} X_i$ où $X_i$ est l'issue du $i$-ème lancer. Les $X_i$ sont mutuellement indépendants et de loi $\mathcal{B}\text{er}(p)$.

4. **Loi géométrique de paramètre $p$ :**

$$x \sim \mathcal{G}\text{eo}(p) \text{ si } \mathbf{P}(X = n) = (1-p)^{n-1} p \text{ pour } n \geq 1.$$

*Interprétation :* On lance une pièce jusqu'à obtenir face. Quelle est la probabilité de lancer la pièce $n$ fois exactement ?

**Proposition 5.** *La loi géométrique est* sans mémoire *: soit $X$ une v.a. de loi géométrique. Alors pour tous $k \geq 0$ et $n \geq 1$, $\mathbf{P}(X = n + k \mid X > k) = \mathbf{P}(X = n)$.*

*Démonstration.*

$$\begin{aligned} \mathbf{P}(X = n + k \mid X > k) &= \frac{\mathbf{P}(X = n + k,\ X > k)}{\mathbf{P}(X > k)} = \frac{\mathbf{P}(X = n + k)}{\mathbf{P}(X > k)} = \frac{(1-p)^{n+k-1} p}{\sum_{i=k}^{\infty} (1-p)^i p} \\ &= \frac{(1-p)^{n+k-1} p}{(1-p)^k} = (1-p)^{n-1} p = \mathbf{P}(X = n). \end{aligned}$$

□

5. **Loi uniforme sur [0,1]** (exception : v.a. non discrète). $\forall x \leq 1$, $\mathbf{P}(X \leq x) = x$.

### 1.2.2 Espérance

**Définition 8.** *Soit $X$ une v.a. à valeurs dans $E$ de distribution $(p(x),\ x \in E)$. Soit $f : E \to \overline{\mathbb{R}}$, où $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. L'espérance de $f(X)$, notée $\mathbf{E}[f(X)]$ est :*

*(a) si $\forall x \in E,\ f(x) \geq 0$, alors $\mathbf{E}[f(X)] = \sum_{x \in E} f(x) p(x)$*

*(b) sinon, soit $f^+(x) = \max(f(x), 0)$ et $f^-(x) = \max(-f(x), 0)$*

1. *si $\mathbf{E}[|f(X)|] < \infty$ (f est intégrable), alors $\mathbf{E}[f(X)] = \mathbf{E}[f(X)^+] - \mathbf{E}[f^-(X)]$.*
2. *si $\mathbf{E}[|f(X)|] = \infty$ et $\mathbf{E}[f(X)^+]$ ou $\mathbf{E}[f^-(X)] < \infty$ (f est sommable), alors $\mathbf{E}[f(X)] = \mathbf{E}[f(X)^+] - \mathbf{E}[f^-(X)] = \pm\infty$.*
3. *sinon $\mathbf{E}[f(X)]$ n'existe pas / n'est pas définie.*

Si $X$ est à valeurs dans $E \subseteq \overline{\mathbb{R}}$, alors $\mathbf{E}[X]$ est sa moyenne.

**Théorème 5** (Linéarité)**.** *Soient $X$ et $Y$ des variables aléatoires réelles d'espérance finie et $a$ et $b$ des réels. Alors*

$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y].$$

**Propriétés importantes**

*Démonstration.* On montre ceci en deux étapes pour plus de clarté. Tout d'abord,

$$\mathbf{E}[aX] = \sum_{x\in E} axp(x) = a\sum_{x\in E} xp(x) = a\mathbf{E}[X].$$

Puis,

$$\begin{aligned}
\mathbf{E}[X+Y] &= \sum_{x,y\in E} (x+y)\mathbf{P}(X=x, Y=y) \\
&= \sum_{x\in E}\sum_{y\in E} x\mathbf{P}(X=x,Y=y) + \sum_{y\in E}\sum_{x\in E} y\mathbf{P}(X=x,Y=y) \\
&= \sum_{x\in E} x\sum_{y\in E} \mathbf{P}(X=x,Y=y) + \sum_{y\in E} y\sum_{x\in E} \mathbf{P}(X=x,Y=y) \\
&= \sum_{x\in E} x\mathbf{P}(X=x) + \sum_{y\in E} y\mathbf{P}(Y=y) \\
&= \mathbf{E}[X] + \mathbf{E}[Y].
\end{aligned}$$

□

**Remarque 2.** *Ce théorème est aussi valide, par le même raisonnement si $X$ et $Y$ sont positives ainsi que $a$ et $b$ dans le cas où l'une de ces v.a. au moins est d'espérance infinie.*

Soient $X$ une v.a. et $\mathcal{P}$ une propriété. Si $\mathbf{P}(X \text{ vérifie } \mathcal{P}) = 1$, on dit que $\mathcal{P}$ est vérifiée presque sûrement (p.s.).

**Proposition 6** (Monotonie)**.** *Soient $X$ une v.a. et $f, g : E \to \overline{\mathbb{R}}$ deux fonctions telles que les espérances de $f(X)$ et $g(X)$ existent. Alors, si $f(X) \leq g(X)$ p.s., alors $\mathbf{E}[f(X)] \leq \mathbf{E}[g(X)]$.*

*Démonstration.* $\mathbf{E}[f(X)] = \sum_{x\in E} f(x)p(x) \leq \sum_{x\in E} g(x)p(x) = \mathbf{E}[g(X)]$. □

**Inégalité de Jensen** pour les fonctions convexes

**Théorème 6** (Inégalité de Jensen)**.** *Si $\phi$ est une fonction convexe sur un intervalle $I \subseteq \mathbb{R}$ et $X$ une v.a. sur $I$, alors si $X$ et $\phi(X)$ sont intégrables, alors*

$$\mathbf{E}[\phi(X)] \geq \phi(\mathbf{E}[X]).$$

*Démonstration.* Si $\phi$ est convexe sur $I$, alors pour tout $x_0 \in \overset{\circ}{I}$, il existe $\alpha$ tel que pour tout $x \in I$, $\phi(x) \geq \phi(x_0) + \alpha(x - x_0)$. Deux cas se présentent : si $X$ est constante presque sûrement, on a $\phi(\mathbf{E}[X]) = \mathbf{E}[\phi(X)]$. Sinon, $\mathbf{E}[X] \in \overset{\circ}{I}$ et on peut poser $x_0 = \mathbf{E}[X]$. Alors, par monotonie, on a

$$\phi(X) \geq \phi(\mathbf{E}[X]) + \alpha(X - \mathbf{E}[X])$$

et en prenant l'espérance, par linéarité,

$$\mathbf{E}[\phi(X)] \geq \mathbf{E}[\phi(\mathbf{E}[X])] + \alpha\mathbf{E}[X] - \alpha\mathbf{E}[\mathbf{E}[X]] = \phi(\mathbf{E}[X]).$$

□

**Espérance conditionnelle par rapport à un événement** Soit $X$ une v.a. et $A$ un événement. L'espérance conditionnelle de $X$ par rapport à $A$ est

$$\mathbf{E}[X \mid A] = \sum_{x \in E} xp(X = x \mid A).$$

**Lemme 1.** *Soient $X$ et $Y$ des variables aléatoires réelles telle que $\mathbf{E}[X]$ existe. Alors*

$$\mathbf{E}[X] = \sum_{y \in E} \mathbf{P}(Y = y)\mathbf{E}[X \mid Y = y].$$

*Démonstration.*

$$\begin{aligned}\sum_{y \in E} \mathbf{P}(Y = y)\mathbf{E}[X \mid Y = y] &= \sum_{y \in E} \mathbf{P}(Y = y) \sum_{x \in E} x\mathbf{P}(X = x \mid Y = y) \\ &= \sum_{x \in E} x \sum_{y \in E} \mathbf{P}(X = x, Y = y) = \sum_{x \in E} x\mathbf{P}(X = x) = \mathbf{E}[X].\end{aligned}$$

□

**Exemples**

1. Si $X$ est constante égale à $a$, alors $\mathbf{E}[X] = a$.
2. Si $X \sim \mathcal{B}\text{er}(p)$, alors
$$\mathbf{E}[X] = 0 \times (1 - p) + 1 \times p = p.$$
Si $A$ est un événement, on utilise souvent la réécriture $\mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$.
3. Si $X \sim \mathcal{B}\text{in}(n, p)$, alors
$$\mathbf{E}[X] = \mathbf{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np,$$
avec $X_i \sim \mathcal{B}\text{er}(p)$.
4. Si $X \sim \mathcal{G}\text{eo}(p)$, alors
$$\mathbf{E}[X] = \sum_{i=1}^{\infty} ip(1 - p)^{i-1} = 1/p.$$
On peut faire le calcul direct ou utiliser la propriété suivante :

**Proposition 7.** *Soit $X$ une v.a. à valeurs dans $\mathbb{N}$. Alors $\mathbf{E}[X] = \sum_{i=1}^{\infty} \mathbf{P}(X \geq i)$.*

*Démonstration.*

$$\sum_{i=1}^{\infty} \mathbf{P}(X \geq i) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \mathbf{P}(X = j) = \sum_{j=1}^{\infty} \sum_{i=1}^{j} \mathbf{P}(X = j) = \sum_{j=1}^{\infty} j\mathbf{P}(X = j) = \mathbf{E}[X].$$

□

Si $X \sim \mathcal{G}\text{eo}(p)$, alors $\mathbf{P}(X \geq i) = \sum_{j=i}^{\infty} \mathbf{P}(X = j) = (1 - p)^{i-1}$ et donc $\mathbf{E}[X] = \sum_{i=1}^{\infty} (1 - p)^{i-1} = 1/p$.

### 1.2.3 Variance

**Définition 9.** *Soit $X$ une variable aléatoire réelle.*
- *Le* $k$-ème moment *d'une v.a.r. $X$ est* $\mathbf{E}[X^k]$.
- *La* variance *d'une v.a.r. $X$ est* $\mathbf{V}\mathrm{ar}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$
- *la* covariance *de deux v.a.r. $X$ et $Y$ est* $\mathbf{C}\mathrm{ov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$.
- *l'*écart-type *de $X$ est* $\sigma(X) = \sqrt{\mathbf{V}\mathrm{ar}(X)}$.

Justification des deux expressions de la variance : $\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2\mathbf{E}[X]X + \mathbf{E}[X]^2] = \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2$

La variance et l'écart-type donnent une mesure de la différence entre $X$ et $\mathbf{E}[X]$. Si la variance est grande, alors avec une forte probabilité, $X$ est très différent de $\mathbf{E}[X]$. Au contraire, si la variance est petite, alors la probabilité que $X$ soit proche de $\mathbf{E}[X]$ est élevée.

**Exemple (variance).** Soit $X$ une v.a. telle que $\mathbf{P}(X = 0) = 1 - 1/k$ et $\mathbf{P}(X = k) = 1/k$. Alors $\mathbf{E}[X] = 1$ et $\mathbf{V}\mathrm{ar}(X) = k - 1$ : $\mathbf{E}[X^2] = k^2/k = k$.

**Proposition 8.** *Soient $X$ et $Y$ deux v.a.r. Alors*
- $\mathbf{V}\mathrm{ar}(X + Y) = \mathbf{V}\mathrm{ar}(X) + \mathbf{V}\mathrm{ar}(Y) + 2\mathbf{C}\mathrm{ov}(X, Y)$
- *Si $X$ et $Y$ sont indépendantes, alors* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.
- *Si $X$ et $Y$ sont indépendantes, alors* $\mathbf{C}\mathrm{ov}(X, Y) = 0$ *et* $\mathbf{V}\mathrm{ar}(X + Y) = \mathbf{V}\mathrm{ar}(X) + \mathbf{V}\mathrm{ar}(Y)$.

*Démonstration.*

$$\begin{aligned}\mathbf{E}[(X + Y - \mathbf{E}[X + Y])^2] &= \mathbf{E}[(X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2 + (Y - \mathbf{E}[Y])^2 + 2(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2] + \mathbf{E}[(Y - \mathbf{E}[Y])^2] + 2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{V}\mathrm{ar}(X) + \mathbf{V}\mathrm{ar}(Y) + 2\mathbf{C}\mathrm{ov}(X, Y).\end{aligned}$$

Si $X$ et $Y$ sont indépendantes, alors

$$\begin{aligned}\mathbf{E}[XY] &= \sum_i \sum_j (i \cdot j)\mathbf{P}(X = i, Y = j) = \sum_i \sum_j (i \cdot j)\mathbf{P}(X = i)\mathbf{P}(Y = j) \\ &= \sum_i i\mathbf{P}(X = i) \sum_j j\mathbf{P}(Y = j) = \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

Mais alors, $\mathbf{C}\mathrm{ov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[(X - \mathbf{E}[X])]\mathbf{E}[(Y - \mathbf{E}[Y])] = 0$ et $\mathbf{V}\mathrm{ar}(X + Y) = \mathbf{V}\mathrm{ar}(X) + \mathbf{V}\mathrm{ar}(Y)$. □

**Exemples**
- Si $X$ est constante égale à $a$, alors $\mathbf{V}\mathrm{ar}(X) = 0$ ;
- Si $X \sim \mathcal{B}\mathrm{er}(p)$, alors $X^2 = X$ et $\mathbf{E}[X^2] = p$. Alors $\mathbf{V}\mathrm{ar}(X) = p - p^2 = p(1 - p)$.
- Si $X \sim \mathcal{B}\mathrm{in}(n, p)$, alors $X = \sum_{i=1}^n X_i$ avec $X_i \sim \mathcal{B}\mathrm{er}(p)$ i.i.d et donc $\mathbf{V}\mathrm{ar}(X) = np(1 - p)$.
- Si $X \sim \mathcal{G}\mathrm{eo}(p)$, alors $\mathbf{V}\mathrm{ar}(X) = \frac{1-p}{p^2}$. En effet, soit $X_0 \sim \mathcal{B}\mathrm{er}(p)$ et $Y \sim \mathcal{G}\mathrm{eo}(p)$ deux v.a. indépendantes. La v.a. $X$ peut s'interpréter comme $X = 1$ si $X_0 = 1$ et $X = Y + 1$ sinon. $\mathbf{E}[X^2] = \mathbf{E}[X^2 \mid X_0 = 0]\mathbf{P}(X_0 = 0) + \mathbf{E}[X^2 \mid X_0 = 1]\mathbf{P}(X_0 = 1) = (1 - p)\mathbf{E}[(Y + 1)^2] + p$. Or $X$ et $Y$ ont même distribution, donc

$$\mathbf{E}[X^2] = (1 - p)(\mathbf{E}[Y^2] + 2\mathbf{E}[Y] + 1) + p = (1 - p)\mathbf{E}[X^2] + 2\frac{1 - p}{p} + 1.$$

Finalement on trouve $\mathbf{E}[X^2] = \frac{2-p}{p^2}$, d'où $\mathbf{V}\mathrm{ar}[X] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

### 1.2.4 Déviations : premières inégalités

Le but des inégalités que nous allons voir dans ce paragraphe est de borner la probabilité que l'écart à la moyenne soit plus grand qu'une certaine valeur. Par exemple, cela pourra être pour un algorithme probabiliste la probabilité que le temps d'exécution soit trop long, pour le collecteur de coupons que le nombre de boîtes à acheter soit trop grand...). On borne plus précisément la quantité $\mathbf{P}(X \geq a)$.

**Théorème 7** (Inégalité de Markov)**.** *Soit $X$ une v.a. réelle positive ou nulle. Alors pour tout $a > 0$,*

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}.$$

*Démonstration.* Soit $a > 0$. On pose $I = 1$ si $X \geq a$ et $I = 0$ sinon. Alors $I \leq \frac{X}{a}$, puisque $a \geq 0$. Alors, $\mathbf{E}[I] = \mathbf{P}(I = 1) = \mathbf{P}(X \geq a)$ et $\mathbf{E}[I] \leq \mathbf{E}[X/a] = \frac{\mathbf{E}[X]}{a}$ (monotonie). □

**Exemple (lancer de pièces).** On lance une pièce non faussée $n$ fois. Borner la probabilité d'obtenir au moins $3n/4$ fois face. On pose $X = \sum_{i=1}^{n} X_i$ où $X_i = 1$ si le $i$-ème lancer est face et $X_i = 0$ s'il est pile. Alors, $\mathbf{E}[X_i] = \mathbf{P}(X_i = 1) = 1/2$. Donc $\mathbf{E}[X] = n/2$. En appliquant l'inégalité de Markov, on obtient $\mathbf{P}(X \geq 3n/4) \leq \frac{n/2}{3n/4} = \frac{2}{3}$.

C'est la borne la moins précise, mais qui est à la base de toutes les autres. On a la même borne pour toutes le v.a. ayant la même espérance et donc cette borne ne dépend pas ici de $n$. Intuitivement, $\mathbf{P}(X \geq 3n/4)$ est décroissante en $n$.

**Théorème 8** (Inégalité de Tchebychev)**.** *Soient $X$ une v.a.r et $a > 0$. Alors*

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) \leq \frac{\mathbf{V}\mathrm{ar}(X)}{a^2}.$$

*Démonstration.* On a $\mathbf{P}(|X - \mathbf{E}[X]| \geq a) = \mathbf{P}((X - \mathbf{E}[X])^2 \geq a^2)$. Comme $(X - \mathbf{E}[X])^2 \geq 0$, on peut lui appliquer l'inégalité de Markov :

$$\mathbf{P}((X - \mathbf{E}[X])^2 \geq a^2) \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} = \frac{\mathbf{V}\mathrm{ar}(X)}{a^2}.$$

□

**Exemple (lancer de pièces).** Reprenons le même exemple du lancer de pièces. On obtient cette fois

$$\mathbf{P}(X \geq 3n/4) \leq \mathbf{P}(|X - \mathbf{E}[X]| \geq n/4) = \frac{\mathbf{V}\mathrm{ar}(X)}{(n/4)^2} = \frac{4}{n}.$$

# 2 Algorithmes probabilistes et méthode probabiliste

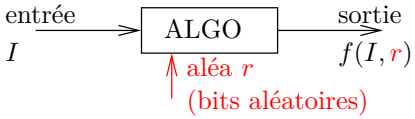
## 2.1 Généralités sur les algorithmes probabilistes

### 2.1.1 Algorithmes probabilistes *v.s.* déterministes

Un algorithme *déterministe* est tel que pour chaque entrée, il existe une et une seule valeur de sortie, qui sera toujours renvoyée.


On veut une réponse rapide et correcte, ce que l'on ne sait pas toujours faire avec un algorithme déterministe. Même un algorithme en $\mathcal{O}(n^3)$ par exemple peut n'être pas assez rapide pour des données de grande taille. Sous réserve que $P \neq NP$, ce n'est pas même pas toujours possible de trouver un algorithme polynomial.

On peut ajouter de l'aléa sous forme de bits aléatoires. Il n'existe donc plus a priori une unique sortie pour chaque entrée.




On peut modifier l'algorithme de sorte à avoir :
— soit une réponse correcte dans la plupart des cas, mais rapide dans tous les cas
— soit une réponse correcte dans tous les cas et rapide dans la plupart des cas.

**Exemple (vérifier l'égalité de deux polynômes).** Considérons deux polynômes de degré $d$, par exemple $F$ et $G$ donnés ci-dessous. Sont-ils égaux ?

$$F(x) = (x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \stackrel{?}{=} x^6 - 7x^3 + 25 = G(x)$$

On peut vérifier cela de plusieurs manières.

**Algorithme 1 :** Algorithme déterministe

**début**
mettre les deux polynômes sous forme canonique ($\sum_{i=0}^{d} c_i x^i$);
vérifier que les coefficients sont égaux.

Un algorithme naïf nécessite $\mathcal{O}(d^2)$ opérations pour développer un polynôme en supposant que les opérations arithmétiques s'effectuent en temps constant (on pourrait aussi faire en $\mathcal{O}(d \log d)$ par « diviser pour régner », mais pour ce premier exemple, on ne compare que des approches naïves).

Si, par exemple, quelqu'un a écrit un programme qui développe un polynôme et veut vérifier son algorithme. Ici, l'algorithme donné passe par le même calcul et nécessite le même temps de calcul. Il présente donc un double inconvénient : le temps de calcul, et la reproduction d'une erreur similaire, en cas d'algorithme erroné.

On peut utiliser l'algorithme aléatoire suivant :
— Si $F = G$, alors l'algorithme renvoie la bonne réponse
— Si $F \neq G$
  — Si $F(r) \neq G(r)$, alors l'algorithme renvoie la bonne réponse

**Algorithme 2 :** Algorithme probabiliste

**début**
  Choisir $r$ dans l'ensemble $\{1, \ldots, 100d\}$ uniformément (chaque nombre a la même chance d'être choisi);
  calculer $F(r)$ et $G(r)$ (en $\mathcal{O}(d)$);
  **si** $F(r) = G(r)$ **alors**
    `Retourner` $F = G$
  **sinon**
    `Retourner` $F \neq G$

— si $F(r) = G(r)$, alors l'algorithme se trompe

L'algorithme se trompe uniquement si $F \neq G$ et $r$ est une racine de $F - G$. Or ce polynôme a au plus $d$ racines dans $\{1, \ldots, 100d\}$. La probabilité que l'algorithme se trompe est alors au plus $1/100$.

### 2.1.2 Classification des algorithmes

**Algorithmes de Las Vegas** [Exemple : algorithme de tri rapide]
— Résout un problème exactement
— avec une complexité moyenne finie (que l'on cherche à minimiser)

**Algorithmes de Monte Carlo** [Exemple : recherche de la médiane]
— Résout un problème de manière approchée (avec une erreur contrôlée)
— avec une complexité qui est une fonction déterministe de la donnée.

Si la probabilité d'erreur d'un algorithme Monte Carlo est $\lambda$, cela signifie que sur chaque entrée, la probabilité d'erreur est au plus $\lambda$, et pas qu'une proportion $\lambda$ de données fournit une réponse erronée.

**Passage d'un algorithme de Monte Carlo à un algorithme Las Vegas** Considérons un algorithme Monte Carlo de complexité $C_{MC}$ qui renvoie soit la réponse correcte, soit la réponse `erreur` (avec probabilité au plus $\lambda$). On peut transformer cet algorithme en un algorithme Las Vegas en le répétant tant que la réponse renvoyée est `erreur`. La complexité de l'algorithme Las Vegas est alors

$$\mathbf{E}(C_{LV}) \leq \frac{C_{MC}}{\lambda}.$$

**Erreur unilatérale** Les algorithmes Monte Carlo qui renvoient `vrai` ou `faux` sont à erreur unilatérale s'ils se trompent seulement sur une réponse, comme c'était le cas pour la vérification de multiplication matricielle ou de polynôme. Pour minimiser l'erreur, il suffit de répéter l'algorithme plusieurs fois.

**Erreur bilatérale** Si un algorithme Monte Carlo qui renvoie les réponses `vrai` ou `faux` peut se tromper sur les deux réponses, il est à erreur bilatérale. Si la probabilité d'erreur est inférieure strictement à $1/2$, alors on peut diminuer cette probabilité en exécutant plusieurs fois l'algorithme et en renvoyant la réponse majoritaire.

**Terminaison**

— Un algorithme termine avec probabilité $\lambda$ si pour toute instance l'algorithme termine avec probabilité au moins $\lambda$.

— Un algorithme termine presque sûrement si pour toute instance il termine avec probabilité 1.

— Un algorithme termine (sûrement) s'il termine sur toutes les entrées, quel que soit le tirage aléatoire.

### 2.1.3 Algorithmes de Las Vegas et principe de Yao

Un algorithme probabiliste de type Las Vegas peut aussi s'interpréter comme une variable aléatoire sur l'ensemble des algorithmes déterministes (résolvant le même problème). Cette interprétation permet de comparer les complexité des algorithmes de Las Vegas et des algorithmes déterministes.

Plus précisément, soit $\mathcal{A}$ un ensemble d'algorithmes déterministes résolvant un certain problème et $\mathcal{X}$ l'ensemble des instances de ces algorithmes. On note $c(a, x)$ le coût (par exemple la complexité) de l'algorithme $a$ sur l'entée $x$.

1. Un algorithme probabiliste $A$ est une variable aléatoire sur $\mathcal{A}$. Le coût moyen de l'algorithme $A$ sur l'entrée $x$ est $\mathbf{E}[c(A, x)]$, et on s'intéresse au coût moyen de l'algorithme sur la *pire* entrée : $\max_{x\in\mathcal{X}} \mathbf{E}[c(A, x)]$.
2. Si on se donne une distribution sur les instances, et $X$ une instance aléatoire distribuée selon cette loi, le coût moyen de l'algorithme déterministe $a$ est $\mathbf{E}[c(a, X)]$, et on s'intéresse au coût moyen du meilleur algorithme déterministe : $\min_{a\in\mathcal{A}} \mathbf{E}[c(a, X)]$.

Le principe de Yao nous indique que la complexité d'un algorithme probabiliste ne peut être meilleur sur toutes les entrées que la complexité moyenne du meilleur algorithme déterministe.

**Théorème 9** (Principe de Yao)**.** *Soit $A$ une variable aléatoire sur $\mathcal{A}$ et $X$ une variable aléatoire sur $\mathcal{X}$. Alors*

$$\max_{x\in\mathcal{X}} \mathbf{E}[c(A, x)] \geq \min_{a\in\mathcal{A}} \mathbf{E}[c(a, X)].$$

*Démonstration.* On a $\mathbf{E}[c(A, x)] = \sum_{a\in\mathcal{A}} \mathbf{P}(A = a)c(a, x)$ et $\mathbf{E}[c(a, X)] = \sum_{x\in\mathcal{X}} \mathbf{P}(X = x)c(a, x)$. Donc

$$\max_{x\in\mathcal{X}} \mathbf{E}[c(A, x)] = \max_{x\in\mathcal{X}} \sum_{a\in\mathcal{A}} \mathbf{P}(A = a)c(a, x) \geq \sum_{x\in\mathcal{X}} \mathbf{P}(X = x) \sum_{a\in\mathcal{A}} \mathbf{P}(A = a)c(a, x)$$
$$\min_{a\in\mathcal{A}} \mathbf{E}[c(a, X)] = \min_{a\in\mathcal{A}} \sum_{x\in\mathcal{X}} \mathbf{P}(X = x)c(a, x) \leq \sum_{a\in\mathcal{A}} \mathbf{P}(A = a) \sum_{x\in\mathcal{X}} \mathbf{P}(X = x)c(a, x).$$

□

## 2.2 La méthode probabiliste

**But :** prouver l'existence d'objets satisfaisant certaines propriétés par des arguments probabilistes. Dans certains cas, on pourra construire effectivement ces objets.

### 2.2.1 Argument de comptage

**Idée :** On dispose d'une collection au plus dénombrable d'objets $a_i$, $i \in I$. On veut prouver que l'un d'eux au moins satisfait une propriété $\mathcal{P}$. Pour ce faire, on peut choisir un objet $a_i$ aléatoirement en introduisant une variable aléatoire $X$ sur $\{a_i\}_{i\in I}$ et si l'on peut prouver que $\mathbf{P}(X \text{ satisfait } \mathcal{P}) > 0$, alors il existe bien $a_i$ qui satisfait $\mathcal{P}$.

**Application : nombre de Ramsey :** coloriage des arêtes d'un graphe complet $K_n$ en deux couleurs, rouge et bleu de telle manière qu'il n'y ait pas de grande clique monochrome. $R(k)$ est la taille minimale du graphe $(n)$ telle qu'il est impossible de trouver un coloriage des arêtes tel qu'il n'y a pas clique de taille $k$ monochrome.

**Théorème 10.** *Si $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, alors $R(k) > n$, c'est-à-dire qu'il est possible de colorier les arêtes de $K_n$ de telle sorte qu'il n'y a pas de clique de taille $k$ monochrome.*

*Démonstration.* Il y a $2^{\binom{n}{2}}$ coloriages possibles des arêtes de $K_n$ avec deux couleurs. On choisit un coloriage uniformément parmi tous ces coloriages possibles. Cela correspond à colorier chaque arête aléatoirement en rouge ou en bleu (chaque couleur avec probabilité 1/2) et indépendamment de la couleur des autres arêtes.

Soit $i = 1, \ldots, \binom{n}{k}$ une énumération des cliques de taille $k$. Soit $A_i$ l'événement « $i$ est une clique monochrome ». Alors

$$\mathbf{P}(A_i) = 2^{-\binom{k}{2}+1} \quad \text{(deux choix parmi } 2^{\binom{k}{2}}\text{)}.$$

Donc

$$\mathbf{P}(\bigcup_{i=1}^{\binom{n}{k}} A_i) \leq \sum_{i=1}^{\binom{n}{k}} \mathbf{P}(A_i) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1.$$

et $\mathbf{P}(\bigcap_{i=1}^{\binom{n}{k}} A_i^c) = 1 - \mathbf{P}(\bigcup_{i=1}^{\binom{n}{k}} A_i) > 0$ et il existe bien un coloriage avec la propriété voulue. □

**Construction effective :** ($k$ est une constante) on utilise un algorithme de type Monte-Carlo.

- Colorier chaque arête uniformément et indépendamment.
- vérifier qu'il n'y a pas de clique de taille $k$ dans un graphe : $\mathcal{O}(n^k)$
- probabilité d'échec : $p = \mathbf{P}(\bigcup_{i=1}^{\binom{n}{k}} A_i)$.
- transformation en un algorithme Las-Vegas : répéter tant qu'on trouve une clique monotone. $\mathbf{E}(\text{temps d'exécution}) = \mathcal{O}\left(\frac{n^k}{1-\binom{n}{k}2^{-\binom{k}{2}+1}}\right)$.

### 2.2.2 Méthode du premier moment (argument de l'espérance)

**Idée :** Si l'espérance d'une v.a. $X$ est $\mu$, alors avec probabilité strictement positive, $X$ prend des valeurs inférieures et supérieures à $\mu$.

**Théorème 11.** *Soit $X$ une v.a.r. Alors $\mathbf{P}(X \geq \mathbf{E}[X]) > 0$ et $\mathbf{P}(X \leq \mathbf{E}[X]) > 0$.*

*Démonstration.* $\mathbf{E}[X] = \sum_x x\mathbf{P}(X = x)$. Si $\mathbf{P}(X \geq \mathbf{E}[X]) = 0$, alors

$$\mathbf{E}[X] = \sum_{x<\mathbf{E}[X]} x\mathbf{P}(X = x) < \sum_{x<\mathbf{E}[X]} \mathbf{E}[X]\mathbf{P}(X = x) = \mathbf{E}[X],$$

ce qui est absurde.

Le même raisonnement est valide si $\mathbf{P}(X \leq \mathbf{E}[X]) = 0$. □

**Application 1 : MAXSAT** **Problème :** $F$ formule en forme normale conjonctive (CNF) (par exemple, $F = (x_1 \vee x_2 \vee x_3) \wedge (x_2 \vee \neg x_3 \vee x_4) \wedge (x_1 \vee \neg x_2 \vee \neg x_4) \ldots$).
**Question :** quel est le nombre maximal de clauses satisfiables ?

*Problème de décision associé* (NP-complet) :

*Données :* $F, k$.

*Question :* existe-t-il une affectation des variables telles que $k$ clauses au moins sont satisfiables ?

**Théorème 12.** *Soient $F$ une formule à $m$ clauses, $k_i$ le nombre de littéraux de la $i$-ème clause et $k = \min_{i=1}^{m} k_i$. Il existe une affectation des variables qui satisfait au moins $\sum_{i=1}^{m}(1-2^{-k_i}) \geq m(1-2^{-k})$ clauses.*

*Démonstration.* On note $x_1, \ldots, x_n$ les variables de la formule. On affecte à chaque variable `vrai` ou `faux` de manière indépendante et uniforme (si $X_i = \mathbf{1}_{x_i \texttt{ vrai}}$ alors $X_i \sim \mathcal{B}\text{er}(1/2)$).

Soit $E_j$ l'événement « la $j$-ème clause est satisfaite » et $Y_j = \mathbf{1}_{E_j}$. On a $\mathbf{E}[Y_j] = 1 - 2^{-k_j}$. Soit $Y = \sum_{j=1}^{m} Y_j$. On a

$$\mathbf{E}[Y] = \sum_j 1 - 2^{-k_j} \geq \sum_j (1 - 2^{-k}) = m(1 - 2^{-k}).$$

Comme $\mathbf{P}(Y \geq \mathbf{E}[Y]) > 0$, il existe une affectation qui satisfait ce nombre de clauses au moins. □

**Algorithme effectif de construction :** utiliser les espérances conditionnelles par rapport à une affectation déjà en partie construite.

$$\begin{aligned}
\mathbf{E}[Y] &= \mathbf{E}[Y \mid X_1 = 1]\mathbf{P}(X_1 = 1) + \mathbf{E}[Y \mid X_1 = 0]\mathbf{P}(X_1 = 0) \\
&= \frac{1}{2}\left(\mathbf{E}[Y \mid X_1 = 1] + \mathbf{E}[Y \mid X_1 = 0]\right) \\
&\leq \max(\mathbf{E}[Y \mid X_1 = 1], \mathbf{E}[Y \mid X_1 = 0]).
\end{aligned}$$

Soit $\mathbf{E}[Y \mid X_1 = 1]$ ou $\mathbf{E}[Y \mid X_1 = 0]$ est supérieur à $\mathbf{E}[Y]$. On choisit l'affectation qui maximise l'espérance : $x_i$ est `vrai` si $\mathbf{E}[Y \mid X_1 = 1] \geq \mathbf{E}[Y \mid X_1 = 0]$, et on recommence pour l'affectation de $x_2$ et des variables suivantes : si on a fixé les valeurs de $X_1, \ldots, X_k$ à $x_1, \ldots, x_k$ respectivement, on fixe celle de $x_{k+1}$. On a

$$\begin{aligned}
\mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k] &= \mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 1]\mathbf{P}(X_{k+1} = 1) \\
&\quad + \mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 0]\mathbf{P}(X_{k+1} = 0) \\
&\leq \max(\mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 1], \\
&\quad \mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 0])
\end{aligned}$$

et on fixe $X_{k+1}$ à 0 ou 1 selon l'espérance.

**Application 2 : ensembles indépendants** Soit $G = (V, E)$ un graphe. On note $|V| = n$ et $|E| = m$. Un sous-ensemble de sommets $I \subseteq V$ est indépendant si $\forall u, v \in I$, $(u, v) \notin E$ ($I$ est une clique dans $G' = (V, \overline{E})$).

**Théorème 13.** *Soit $G = (V, E)$ un graphe connexe de $n$ sommets et $m$ arêtes. Si $\frac{2m}{n} \geq 1$, alors $G$ possède un ensemble indépendant de taille au moins $n^2/4m$.*

*Démonstration.* Si le graphe contient plus de sommets que d'arêtes, alors on sait qu'il existe un indépendant de taille $n - m$. En effet, on peut enlever un sommet extrémité de chaque arête, et l'on obtient un graphe sans arête d'au moins $n - m$ sommets. Une première phase d'un algorithme de construction consiste donc à enlever des sommets (et les arêtes qui y sont adjacentes) afin de construire un tel graphe.

Soit $p \in [0, 1]$.

**Algorithme 3 :** Trouver un indépendant

**début**

Effacer chaque sommet de $G$ avec probabilité $1 - p$ indépendamment ;
Pour chaque arête restante, la retirer, ainsi qu'un des sommets adjacents.

On peut calculer l'espérance du nombre de sommets et d'arêtes obtenus à la fin de la seconde phase.

1. Soit $X$ le nombre de sommets restant après la première étape : $\mathbf{E}[X] = np$.
2. Soit $Y$ le nombre d'arêtes à la fin de la première étape. Une arête survit si aucun des sommets qu'elle relie n'est supprimé, ce qui arrive avec probabilité $p^2$. Ainsi, $\mathbf{E}[Y] = mp^2$.

A la deuxième étape, on retire un sommet par arête au plus. Le nombre de sommets restants est alors au moins $X - Y$ et $\mathbf{E}[X - Y] = np - mp^2$. Cette quantité est maximisée pour $p = \frac{n}{2m} \leq 1$ par hypothèse. On obtient alors $\mathbf{E}[X - Y] = \frac{n^2}{2m} - \frac{n^2}{4m} = \frac{n^2}{4m}$. Notons que $d = \frac{1}{p} = \frac{2m}{n}$ est le degré moyen du graphe. □

### 2.2.3 Lemme local de Lovász (version symétrique)

Les méthodes étudiées jusqu'à présent utilisent pour la plupart le fait que les événements mis en jeu sont indépendants ou alors utilisent l'union bound.

Si $E_1, \ldots, E_n$ sont des *mauvais* événements et qu'on veut montrer qu'il existe des situations où aucun de ces mauvais événements n'a lieu, il suffit de montrer que

$$\mathbf{P}(\overline{\cup E_i}) > 0.$$

Deux solutions sont possibles :

1. $\mathbf{P}(\overline{\cup E_i}) = 1 - \mathbf{P}(\cup E_i) = 1 - \sum_{i=1}^{n} \mathbf{P}(E_i)$ (Union bound)
2. $\mathbf{P}(\overline{\cup E_i}) = \mathbf{P}(\cap \overline{E_i}) = \prod_{i=1}^{n} \mathbf{P}(\overline{E_i})$ si les $E_i$ sont mutuellement indépendants.

Comment faire si la première méthode ne suffit pas ou si les événements ne sont pas mutuellement indépendants ? Le lemme local de Lovász permet de donner une réponse à cette question lorsque les événements ne sont pas *trop* dépendants.

On dit que qu'un événement $E$ est *mutuellement indépendant* de $E_1, \ldots, E_n$ si pour tout $I \subseteq \{1, \ldots, n\}$, $\mathbf{P}(E \mid \cap_{i \in I} E_i) = \mathbf{P}(E)$.

**Définition 10** (Graphe de dépendance)**.** *Un graphe de dépendance de $E_1, \ldots, E_n$ est un graphe $G = (V, E)$ tel que $V = \{1, \ldots, n\}$ et $E_i$ est mutuellement indépendant des $\{E_j \mid (i, j) \notin E\}$.*

Par exemple, prenons l'exemple d'une formule en CNF :

$$F = (x_1 \vee x_2 \vee \neg x_3) \wedge (x_1 \vee \neg x_2 \vee x_3) \wedge (x_2 \vee x_3 \vee x_4) \wedge (x_4 \vee x_5 \vee \neg x_6).$$

Les mauvais événements sont $E_j$ : « la $j$-ème clause n'est pas satisfaite ». Supposons que les valeurs des variables sont affectées les unes indépendamment des autres. Alors, par exemple, $E_4$ est mutuellement indépendant de $E_1$ et $E_2$ car la 4ème clause ne partage pas de littéraux avec les deux premières.

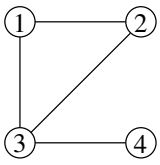



FIGURE 1 – Graphe de dépendance correspondant à $F$.

**Théorème 14** (Lemme local de Lovász symétrique)**.** *Soient $E_1, \ldots, E_n$ des événements tels que*

1. $\forall i \in \{1, \ldots n\}$, $\mathbf{P}(E_i) \leq p$.
2. *Le degré des sommets du graphe de dépendance de $E_1, \ldots, E_n$ est borné par $d$.*
3. $4pd \leq 1$.

*Alors*

$$\mathbf{P}(\cap_{i=1}^{n} \bar{E_i}) > 0.$$

En utilisant l'Union bound, il faudrait que $pn < 1$. Le lemme de Lovász donne de meilleurs résultats dès que $d \leq n/4$.

**Exemple ($k-SAT$). Données :** Une formule $F$ en CNF avec $k$ variables distinctes exactement par clause.
**Question :** $F$ est-elle satisfiable ?

**Théorème 15.** *Soit $F$ une formule en CNF avec $k$ littéraux exactement par clause. Si aucune variable n'apparaît plus de $2^k/4k$ fois ou si aucune clause ne partage de variable en commun avec $2^{k-2}$ autres clauses, alors $F$ est satisfiable*

*Démonstration.* Soit $E_j$ l'événement «  la clause $j$ n'est pas satisfaite ». On affecte les variables indépendamment selon une loi de Bernoulli de paramètre 1/2. Alors $\mathbf{P}(E_j) \leq 2^{-k}$.

$E_j$ est mutuellement indépendant des événements concernant des clauses qui n'ont pas de variable en commun avec la $j$-ème.

Chaque variable apparaît au plus $2^k/4k$ fois, donc $d \leq k \times 2^k/4k = 2^{k-2}$. (On a le même degré avec l'autre énoncé). Alors, $4dp \leq 1$. □

Avec $k = 3$, une clause peut partager d'autres variables avec deux clauses.

## 2.3 Exercices

**Exercice 5** *Analyse du tri rapide*
On rappelle l'algorithme du tri rapide :

**Algorithme 4 :** Tri_Rapide

**Données :** Une liste $S$ de $n$ entiers distinct
**Résultat :** La liste triée des éléments de $S$
**début**
  **si** *$S$ a 0 ou 1 élément* **alors** retourner $S$;
  **sinon**
    Choisir un élément $x$ (pivot) de $S$ et diviser les autres éléments en deux sous-listes
      — $S_1$, liste des éléments de $S$ qui sont $< x$ ;
      — $S_2$, liste des éléments de $S$ qui sont $> x$;
    Tri_Rapide($S1$) ; Tri_rapide($S_2$);
    Retourner la liste $S_1, x, S_2$.

**1.** Donner un exemple de liste qui nécessite $\Omega(n^2)$ comparaisons pour la trier avec cet algorithme.

On veut montrer que si les pivots sont choisis indépendamment et uniformément, alors l'espérance du nombre de comparaisons est de $2n \ln n + O(n)$. On note $y_1 < y_2 < \cdots < y_n$ les éléments de la liste.

**2.** Quelle est la probabilité que deux éléments $y_i$ at $y_j$ soient comparés ?

**3.** En déduire le résultat.

**4.** Que se passe-t-il si on choisit toujours le premier élément de la liste comme pivot ? Quelle est la différence avec le choix d'un pivot aléatoire ?

**Exercice 6** *Recherche de la médiane*

On considère l'algorithme suivant :

**Algorithme 5 :** Calcul de la médiane randomisé

**Données :** Un ensemble $S$ de $n$ entiers
**Résultat :** La médiane de $S$, noté $m$
**début**
  Choisir un multi-ensemble $R$ de $\lceil n^{3/4} \rceil$ éléments de $S$ choisis uniformément et indépendamment;
  Trier $R$;
  Soit $d$ le $\lfloor \frac{1}{2} n^{3/4} - \sqrt{(}n) \rfloor$-ième plus petit élément de $R$;
  Soit $u$ le $\lceil \frac{1}{2} n^{3/4} + \sqrt{(}n) \rceil$-ième plus petit élément de $R$;
  Comparer tous les éléments de $S$ à $d$ et $u$. Soient $C = \{x \in S \mid d \leq x \leq u\}$, $\ell_d = |\{x \in S \mid x < d\}|$ et $\ell_u = |\{x \in S \mid x > u\}|$;
  **si** $\ell_d > n/2$ *ou* $\ell_u > n/2$ **alors** ÉCHEC;
  ;
  **sinon si** $|C| > 4n^{3/4}$ **alors** ÉCHEC;
  ;
  **sinon**
    Trier $C$;
    Retourner le $\lfloor n/2 \rfloor - \ell_d + 1$-ième élément de $C$ trié.

**1.** Expliquer pourquoi cet algorithme termine en $O(n)$ et pourquoi cet algorithme renvoie soit ÉCHEC soit la bonne réponse.

On veut maintenant connaître la probabilité que la réponse soit ÉCHEC. Un ÉCHEC appartient à au moins un de ces trois événements :

— $\mathcal{E}_1 : Y_1 = |\{r \in R \mid r \leq m\}| < \frac{1}{2}n^{3/4} - \sqrt{n}$ ;
— $\mathcal{E}_2 : Y_2 = |\{r \in R \mid r \geq m\}| < \frac{1}{2}n^{3/4} - \sqrt{n}$ ;
— $\mathcal{E}_3 : |C| > 4n^{3/4}$.

**2.** Montrer que $P(\mathcal{E}_1) = P(\mathcal{E}_2) \leq \frac{1}{4}n^{-1/4}$.

**3.** Montrer que $P(\mathcal{E}_3) \leq \frac{1}{2}n^{-1/4}$.

**4.** Conclure.

**Exercice 7** *Coupe d'un graphe*

Soit $G = (V, E)$ un graphe non-orienté. Une coupe de $G$ est une partition des sommets en $A$ et $B$, deux ensembles disjoints. La valeur d'une coupe $C(A, B)$ est le nombre d'arêtes reliant un sommet de $A$ à un sommet de $B$. Le problème de trouver une coupe de valeur maximale est NP-difficile, et nous allons donner un algorithme construisant un coupe de valeur $\geq |E|/2$. On pose $|E| = m$.

**1.** On construit $A$ et $B$ aléatoirement, en affectant indépendamment et uniformément à chaque sommet un des deux ensembles $A$ ou $B$. Quelle est la probabilité qu'une arête appartienne à la coupe ? En déduire qu'il existe une coupe de valeur $\geq m/2$.

**2.** Montrer que

$$P(C(A, B) \geq m/2) \geq \frac{1}{m+1}.$$

Quel est le nombre d'essais à faire avant de trouver une coupe de valeur $\geq m/2$ ? En déduire un algorithme aléatoire probabiliste pour trouver une coupe de valeur $\geq m/2$.

Nous allons maintenant voir comment trouver un algorithme déterministe glouton pour résoudre le même problème.

Supposons que l'on a déjà affecté les sommets $v_1, \ldots, v_k$ à $A$ ou $B$. On associe au placement de $v_i$ l'événement $x_i$.

**3.** Montrer que l'on peut placer le sommet $v_{k+1}$ de telle sorte que

$$E[C(A, B) \mid x_1, \ldots, x_k] \leq E[C(A, B) \mid x_1, \ldots, x_{k+1}].$$

**4.** Conclure.

**Exercice 8** *Coupe minimale dans un graphe*

Soit $G = (V, E)$ un multigraphe (graphe avec des arêtes multiples) sans boucle et connexe. Une *coupe* est un ensemble d'arêtes qui, si elles sont retirées du graphe, le déconnecte. On souhaite touver le cardinal minimal d'une coupe.

On propose un algortihme probabiliste et le but est de calculer une borne sur la probabilité de succès de cet algorithme.

L'algorithme consiste en une succession de contraction d'arêtes à chaque étape, jusqu'à ce qu'il ne reste que deux sommets dans le graphe, on sélectionne une arête uniformément parmi toutes les arêtes du graphe, on identifie les deux extrémités de cette arête et enfin on supprime toutes le boucles créées (c'est-à-dire toutes les arêtes entre ces deux sommets, qui ne forment maintenant plus qu'un sommet). À la fin, on retourne le nombre d'arêtes entre les deux sommets restants.

**1.** Montrer que les arêtes restantes sont initialement une coupe du graphe, et donc que le résultat renvoyé est au moins égal au cardinal de la coupe minimale.

Soit $n$ le nombre de sommets du graphe initial, $m$ le nombre d'arêtes et $k$ le cardinal d'une coupe minimale, et $\mathcal{C}$ une coupe minimale.

**2.** Montrer que $m \geq \frac{kn}{2}$.

On calcule maintenant la probabilité qu'aucune arête de $\mathcal{C}$ ne soit choisie. On note $A_i$ l'événement « aucune arête de $\mathcal{C}$ n'est choisie à l'étape $i$ pour la contraction ».

**3.** Calculer $\mathbf{P}(A_i \mid A_1, \ldots, A_{i-1})$.

**4.** En déduire que la probabité de succès de l'algorithme est au moins de $\frac{2}{n^2}$.

**5.** Donner une borne sur le nombre d'exécutions à effectuer pour avoir une probabilité d'échec d'au plus $1/e$ ?

**Exercice 9** *Lemme local de Lovász symétrique*

Le but de cet exercice est de montrer le théorème suivant :

Soient $E_1, \ldots, E_n$ des événements dans un espace de probabilité arbitraire, et $G = (V, E)$ le graphe de dépendance de ces événements. Si

— $\forall i \in \{1, \ldots n\}$, $\mathbf{P}(E_i) \leq p$ ;
— le degré des sommets du graphe de dépendance est borné par $d$ ;
— $4pd \leq 1$,

alors

$$\mathbf{P}(\cap_{i=1}^{n} \bar{E}_i) > 0.$$

Soit $S \subseteq \{1, \ldots, q\}$. On va montrer par récurrence sur $s \in \{0, \ldots, n-1\}$ que si $|S| \leq s$, alors pour tout $k \notin S$,

$$P(E_k \mid \cap_{j \in S} \bar{E}_j) \leq 2p. \qquad \text{(H)}$$

**1.** Montrer que ceci est vrai pour $s = 0$.

On suppose que la propriété est vraie jusqu'à l'ordre $s$ et on considère un ensemble $S$ de cardinal $s + 1$.

**2.** Montrer que si $|S| \leq n$, $P(\cap_{j \in S} \bar{E}_j) > 0$.

On fixe $k \notin S$, on pose $S_1 = \{j \in S \mid (k, j) \in E\}$ et $S_2 = S - S_1$, ainsi que $F_1 = \cap_{j \in S_1} \bar{E}_j$, $F_2 = \cap_{j \in S_2} \bar{E}_j$ et $F = F_1 \cap F_2$.

**3.** Montrer que (H) est satisfaite si $S_2 = S$.

**4.** Montrer que $P(E_k \mid F) = \frac{P(E_k \cap F_1 \mid F_2)}{P(F_1 \mid F_2)}$.

**5.** Montrer que $P(E_k \cap F_1 \mid F_2) \leq p$ et que $P(F_1 \mid F_2) \geq 1/2$. En déduire que (H) est satisfaite.

**6.** Conclure.

**Exercice 10** *Coloriage*

Soit $G = (V, E)$ un graphe non-orienté et supposons qu'à chaque sommet $v \in V$ est associé un ensemble $S(v)$ de $8r$ couleurs, $r \geq 1$. En outre, pour chaque sommet $v$ et chaque couleur $c \in S(v)$, il y a au plus $r$ voisins $u$ de $v$ tels que $c \in S(u)$.

Montrer qu'il existe alors un coloriage propre (pas deux sommets adjacents avec la même couleur) tel que pour tout $v$, la couleur associée à $v$ est choisie dans $S(v)$.